



ISA White Paper

Fifty Years of Microprocessor Technology Advancements: 1965 to 2015

Table of Contents

Executive Summary	3
Changing the Focus: More Cores Instead of More Transistors	4
Replacing the x86 Paradigm	6
Advancing to Multi-Core Designs	7
Leveraging Software Tools for Optimal Performance	9
Optimizing Microarchitecture through Parallelism.....	9
Combining Performance with Security	11
Further Advancing the Microprocessor: 2015 and Beyond.....	11
Summary Conclusion	12

Executive Summary

Everyone who works in the computer industry is well familiar with Moore's Law and the doubling of the number of transistors (an approximate measure of computer processing power) every 18 to 24 months. Until recently, overall microprocessor performance was often described in terms of processor clock speeds, expressed in megahertz (MHz) or gigahertz (GHz).

Today there's far more than clock speed to consider when you're evaluating how a given processor will perform for a given application and where it fits on the performance scale. Microprocessor designers today are more focused on methods that leverage the latest silicon production processes and designs that minimize microprocessor footprint size, power consumption and heat generation.

Designers are also concerned with microarchitecture optimization, multi-processing parallelism, reliability, designed-in security features, memory structure efficiency and better synergy between the hardware and accompanying software tools, such as compilers. The more attention that a designer devotes to refining the efficiency of the software code rather than making the hardware responsible for dynamic optimization, the higher the ultimate system performance will be.

As an example, the Intel® Itanium® processor family has been designed around small footprint cores that are remarkably compact in terms of transistor count, especially when one considers the amount of processing work that they accomplish. Itanium has taken instruction level parallelism to a new level, and this can be used in conjunction with thread level parallelism to leverage more processor cores and more threads per core to produce higher performance.

Some microprocessor designs of the past have been overly complex and have relied on out-of-order logic to reshuffle and optimize software instructions. Going forward, microprocessor designers will continue to deliver better and better software tools, higher software optimization and better compilers.

Because it is so efficient and so small and doesn't depend on out-of-order logic, the latest generation Itanium processor can deliver higher performance without creating thermal generation problems. This makes Itanium a very simple yet efficient and refined engine that enables more consistent long-term improvement in code execution via small improvements in software, thus reducing the need for significant advancements in hardware. These are becoming more and more difficult to accomplish as, even Gordon Moore believes, the exponential upward curve in microprocessor hardware advancements "can't continue forever."

Changing the Focus: More Cores Instead of More Transistors

Microprocessor advancement accelerated rapidly in 1968 when three engineers from Fairchild Semiconductor – Robert Noyce, Andy Grove and Gordon Moore – founded Intel Corporation in Mountain View, California, to develop new technologies for silicon-based chips.

In 1971, Intel developers successfully embedded a central processing unit, memory, input and output controls on the world's first single chip microprocessor, the Intel 4004 (U.S. Patent #3,821,715) (“Inventors of the Modern Computer,” by Mary Bellis, About.com, undated).

Moore was well known in the industry for a 1965 article where he predicted the doubling of the number of transistors on microprocessors (an approximate measure of processing power) every 18 to 24 months. For more than 40 years, what became known as Moore's Law has accurately described the pace of advancements that have driven the semiconductor industry to reach over \$200 billion in annual revenue today and serve as the foundation of the trillion-dollar electronics industry. (“Moore Optimistic on Moore's Law,” Intel Corporation, <http://www.intel.com/technology/silicon/mooreslaw/eml02031.htm>)

But even Gordon Moore, as quoted recently by Techworld, says Moore's Law “can't continue forever.” (“Moore's Law is dead, says Gordon Moore” by Manek Dubash, Techworld, April 13, 2005). Heat generation and circuit power requirements become more and more of a barrier as transistor density and processor clock speeds increase.

Either directly or indirectly, processor clock speed, expressed in Megahertz (MHz) or Gigahertz (GHz), was once the common reference point used to predict how a given processor would perform for a given application and where a processor ranked on performance comparisons. However, microprocessor designers today are more focused on other properties that enable processors to deliver higher performance. Among these are methods that leverage the latest silicon production processes and designs that minimize microprocessor footprint size, power consumption and heat generation.

Designers are also concerned with microarchitecture optimization, multi-processing parallelism, reliability, designed-in security features, memory structure efficiency and better synergy between the hardware and accompanying software tools, such as compilers and system and applications libraries. The more attention that a designer devotes to refining the efficiency of the software code rather than making the hardware responsible for dynamic optimization, the higher the ultimate system performance will be.

Widely available microprocessors in 1993 had around three million transistors while the Intel® Itanium® processor currently has nearly one billion transistors. If this rate continued,” writes science and technology journalist Geoff Koch, “Intel processors would soon be producing more heat per square centimeter than the surface of the sun—which is why the problem of heat is already setting hard limits to frequency (clock speed) increases.” (“Discovering Multi-Core: Extending the Benefits of Moore's Law,” by Geoff Koch, Technology@Intel Magazine (online), undated)

Increasing processor performance without producing excessive heat is a challenge that can be solved, in part, by dual- and multi-core processor architecture, according to researchers at IDC (“The Next Evolution in Enterprise Computing,” by Kelly Quinn, Jessica Yang and Vernon Turner, IDC, April 2005).

Multi-core chips produce higher performance without a proportionate increase in power consumption and only a minimal increase in heat generation. By increasing the number of cores rather than the number of transistors on a single core, performance advancements can continue indefinitely by leveraging the operational benefits of microprocessor parallelism and process concurrency.

“If we were to continue down the Gigahertz path, the power requirements and heat problems of processors would get out of hand,” says Paul Barr, a technology manager at Intel. Looking ahead over the next three to four years, Barr expects to see as much as a 10x boost in microprocessor performance due to multi-core processors and multi-threaded applications. “Multi-core is the next generation,” Barr explains. “It’s just a natural progression of Moore’s law.” (Paul Barr, Intel, interviewed by the Itanium Solutions Alliance.).

Today there's far more than clock speed to consider when you're evaluating how a given processor will perform for a given application and where it fits on the performance scale.

“There is no doubt that the whole industry has shifted the focus away from ramping clock speed and improving ILP (instruction level parallelism) to increasing performance by exploiting TLP (thread level parallelism),” says popular European technology writer Johan De Gelas. (“Itanium - is there light at the end of the tunnel?” by Johan De Gelas, Ace’s Hardware, Nov 9, 2005)

Writing in Dr. Dobb's Journal, developer Herb Sutter estimates that "Like all exponential progressions, Moore's Law must end someday, but it does not seem to be in danger for a few more years yet. Despite the wall that chip engineers have hit in juicing up raw clock cycles, transistor counts continue to explode and it seems CPUs will continue to follow Moore's Law-like throughput gains for some years to come." ("The Free Lunch Is Over: A Fundamental Turn Toward Concurrency in Software," by Herb Sutter, Dr. Dobb's Journal, March 2005.

"As you move to multiple-core devices, scaling the frequency higher isn't as important as the ability to put multiple cores on a chip," says Dean McCarron, principal analyst at Mercury Research Inc., quoted in eWeek. (Intel Swaps Clock Speed for Power Efficiency, by John G. Spooner eWeek, August 15, 2005)

This is not to say that single-core clock speeds won't increase – they will – but future advancements will happen at a slower pace. At the same time, dual-core processors are expected to offer substantial performance improvements over single core by running a little faster and taking full advantage of dual core benefits such as parallelism and other things such as support for the Message Passing Interface (MPI), a standardized API typically used for parallel and/or distributed computing, created by the MPI Forum (www.mpi-forum.org).

"The immutable laws of physics don't necessarily lead to hard limits for computer users," Geoff Koch adds. "New chip architectures built for scaling out instead of scaling up will offer enhanced performance, reduced power consumption and more efficient simultaneous processing of multiple tasks." ("Discovering Multi-Core: Extending the Benefits of Moore's Law," by Geoff Koch, Technology@Intel Magazine (online), undated).

Replacing the x86 Paradigm

As Thomas Kuhn once observed ("The Structure of Scientific Revolution," Thomas Kuhn, 1962.), scientific advancement is not evolutionary, but is rather a "series of peaceful interludes punctuated by intellectually violent revolutions", and in those revolutions "one conceptual world view is replaced by another."

The conceptual framework of the Intel x86 or 80x86 microprocessor architecture was first introduced by Intel in 1978, and has since followed an unwavering path of advancement. Starting with an 8-bit instruction set, the x86 ISA grew to a 16-bit and then to a 32-bit instruction set. These labels (8-bit, 16-bit, 32-bit) designate the number of bits that each of the microprocessor's general-purpose registers (GPRs) can hold. The term "32-bit processor" translates to "a processor with GPRs that store 32-bit numbers." Similarly, a "32-bit instruction" is an instruction that operates on 32-bit numbers.

A 32-bit address space allows the CPU to directly address 4 GB of data. Though 4 GB once seemed gargantuan, the size requirements of memory-intensive applications such as multimedia programs or database query engines are often much higher. In response to this shortcoming, the 64-bit microarchitecture has increased RAM addressability from 4 GB to a theoretical 18 million terabytes (10¹⁹ bytes). However, since the virtual address space of x86-64 is 48 bits and the physical address space is 40 bits, in reality the yield is only an approximate 256 terabytes. The 64-bit processor also has registers and arithmetic logic units that can manipulate larger instructions (64-bits worth) at each processing step.

Since 64-bit processors can handle chunks of data and instructions twice as large as 32-bit processors, the 64-bit microarchitecture should theoretically be able to process twice as much data per clock cycle as the 32-bit microprocessor.

Unfortunately, that's not quite true. As Jonathan Stokes writes, "only applications that require and use 64-bit integers will see a performance increase on 64-bit hardware that is due solely to a 64-bit processor's wider registers and increased dynamic range." ("An Introduction to 64-bit Computing and x64," by Jonathan Stokes, arstechnica.com, undated.)

Advancing to Multi-Core Designs

In order to advance beyond current x86 capabilities without adding more transistors, microprocessor designs are now headed down two paths. One path represents an embellishment of the traditional x86 design – multiple cores packaged on one board running parallel threads on each.

The other path – Intel Itanium – represents one of those paradigm shifts described by Kuhn as a "revolution in which one conceptual world view is replaced by another."

Conceptually, multi-core architecture refers to a single processor package containing two or more processor "execution cores," or computational engines that deliver fully parallel execution of multiple software threads. The operating system treats each of its execution cores as a discrete processor, with all associated execution resources.

Multi-core processors thus deliver higher performance and greater efficiency without the heat problems and other disadvantages experienced by single core processors run at higher frequencies to squeeze out more performance. By multiplying the number of cores in the processor, it is possible to dramatically increase computing resources, higher multithreaded throughput, and the benefits of parallel computing. ("Intel Multi-core Platforms, Intel Corporation, www.intel.com/technology/computing/multi-core/index.htm, undated)

Although multi-core technology was first discussed by Intel in 1989 ("Microprocessors Circa 2000" by Intel VP Pat Gelsinger and others, IEEE Spectrum, October 1989.), the company released its first dual-core processor in April 2005 to mark the first step in its transition to multi-core computing. The company is now engaged in research on architectures that could include dozens or even hundreds of processors on a single die. ("Intel Multi-core Platforms, Intel Corporation, www.intel.com/technology/computing/multi-core/index.htm, undated)

Intel has publicly committed itself to a vision of "moving beyond gigahertz" to deliver greater value, performance and functionality with multi-core with multi-core architectures and a platform-centric approach.

Central to this strategy, Intel and its industry partners allied through the Itanium Solutions Alliance (founded in September, 2005) are making billions of dollars in strategic technology investments to assure that the Intel Itanium processor becomes the platform of choice for mission critical enterprise systems and technical computing within four years. ISA founding sponsors include Bull, Fujitsu, Fujitsu Siemens Computers, Hitachi, HP, Intel, NEC, SGI and Unisys. Charter members include BEA, Microsoft, Novell, Oracle, Red Hat, SAP, SAS and Sybase. Over a dozen additional technology organizations have also joined the alliance.

According to Lisa Graff, general manager of Intel's high-end server group, Itanium is already being well accepted in the mission-critical systems marketplace, with half of the world's 100 largest enterprises now deploying the Itanium platform. Graff told CNET News "I think Itanium is the architecture for the next 20 years. It's the newest architecture that has come out. It has the headroom. I think the RISC architectures will run out of steam." ("Itanium: A cautionary tale," By Stephen Shankland, CNET News.com, December 7, 2005).

Researchers at Gartner, Inc., quoted by CNET, estimate that the Itanium-based server market is currently about 42,000 units, or about \$2.6 billion in sales. By 2010, Itanium is projected by Gartner to expand 234,000 servers, or about \$7.7 billion in market value.

Leveraging Software Tools for Optimal Performance

Some microprocessor designs of the past have been overly complex and have relied on out-of-order logic to reshuffle and optimize software instructions. Going forward, designers will continue to deliver better and better software tools, higher software optimization and better compilers.

Among semiconductor companies, Intel provides one of the strongest suites of software tools to support and enhance the performance of microprocessors such as the Intel Itanium. This includes an 18-month roadmap that Intel shares with its major customers showing how its software tools will advance and evolve over time. This enables Itanium developers to deliver higher performance to their users and customers.

Among the most important software tools from Intel are a family of compilers optimized for each of its microprocessor platforms, including Xeon and Itanium. According to Intel, some of these compilers work cross platform, often enabling one set of source code to be delivered for multiple platforms, occasionally requiring a simple recompile for each platform. (John McHugh, Intel, interviewed by the Itanium Solutions Alliance.).

Leveraging a compiler to enhance code performance for Itanium applications is a key benefit for Itanium users, as explained by technology writer Johan De Gelas. "The main philosophy behind Itanium is... that a compiler can statically schedule instructions much better than a hardware scheduler, which has to decide this dynamically in a few clock cycles... the compiler can search through thousands of instructions ahead while the hardware scheduler can check only a few tens of instructions for independent instructions. The compiler will make groups of instructions that can be issued simultaneously without dependencies or interlocks. These groups can be one or tens of instructions." ("Itanium - is there light at the end of the tunnel?" by Johan De Gelas, Ace's Hardware, Nov 9, 2005)

In addition to Itanium compilers, Intel offers a math kernel library (MKL) and an integrative performance primitives library (IPP), that can significantly enhance Itanium code performance for scientific cluster applications, video image processing, audio processing and image recognition.

Optimizing Microarchitecture through Parallelism

As microprocessor technology continues to advance with new architectures such as the Intel Itanium, efficiency and performance optimization become even more critical. A new technique used to achieve this is multi-processing parallelism and a shift from instruction level parallelism to thread level parallelism.

The Intel Itanium processor family is designed around small footprint cores that are remarkably compact in terms of transistor count, especially when one considers the amount of processing work that they accomplish. Itanium has taken instruction level parallelism to a new level, and this can be used to leverage more processor cores and more threads per core to produce higher performance.

When Intel introduced Itanium in 2001, the company made a commitment to a design that takes a quantum leap ahead of instruction level parallelism. Instruction level parallelism is a process in which independent instructions (instructions not dependent on the outcome of one another) execute concurrently to utilize more of the available resources of a processor core and increase instruction throughput. The ability of the processor to work on more than one instruction at a time improves the cycles per instruction and increases the frequency from previous architectures. (“A Recent History of Intel Architecture,” by Sara Sarmiento, Intel Corporation, undated.)

Contrast this with a processor equipped with thread-level parallelism that executes separate threads of code. This could be one thread running from an application and a second thread running from an operating system, or parallel threads running from within a single application.

In the past, microprocessor advancements have been based on improving the performance of a single thread. Since there was only one core per processor, advancements were achieved by adding more transistors and increasing the speed of each transistor to improve program performance. This, along with various pipelining and code strategies, made it possible for multiple instructions to be issued in parallel. (John Crawford, Intel, interviewed by the Itanium Solutions Alliance.).

Moving forward, Itanium will leverage multiple cores and thread level parallelism to produce greater performance enhancements that would be possible through single cores and single thread performance boost.

This move toward chip-level multiprocessing architectures with a large number of cores continues a decades-long trend at Intel, offering dramatically increased performance and power characteristics. (“Platform 2015 Software: Enabling Innovation in Parallelism for the Next Decade,” by David J. Kuck, Technology@Intel Magazine, (online) undated). This also presents significant challenges, including a need to make multi-core processors easy to program, which is accomplished, in part, using the software tools described above.

Combining Performance with Security

The dramatic performance advancements delivered by Itanium make it possible for additional security features to be incorporated into applications, enabling developers to avoid the trade-offs often made between performance and security.

Itanium provides four privilege levels that the operating system can leverage to provide a clean separation between what the user can access versus what the virtual machine can access. More importantly, Itanium provides a protection key scheme based on container logic. With the appropriate application support, a developer can compartmentalize everything contained in vast memory stores, and, in effect, put a security wrapper around each process. (Dave Myron, Intel, interviewed by the Itanium Solutions Alliance.)

Itanium's ability to run high degrees of instructions per cycle means that developers can create security layers that run very efficiently to protect code stacks against a variety of attacks. There's also an application that takes strong advantage of the instructions per cycle to create an efficient security layer. In addition to that, there are floating point units unique to Itanium that provide very fast encryption.

The combination of all of these features – parallelism, floating point units, privilege levels, and the protection key scheme, when taken together and using the right application, provide microarchitecture security that is second-to-none.

Further Advancing the Microprocessor: 2015 and Beyond

Considering the dramatic progress made in microprocessor design and architecture since 1965, it's risky to project what new technologies could become available in ten to fifteen years. And yet, a group of future-minded researchers are expressing optimism about the potential of tiny nanoelectronic components, organic molecules, carbon nanotubes and individual electrons that could serve as the underlying technology for new generation of microprocessors emerging around 2015.

There are limits to what can be accomplished with silicon, says Philip J. Kuekes, a physics researcher at Hewlett-Packard Laboratories quoted in the New York Times ("Chip Industry Sets a Plan For Life After Silicon," by John Markoff, New York Times, December 29, 2005). Kuekes says H-P is currently working on molecular-scale nanotechnology switches that, it is hoped, will be able to overcome some of the present-day technological challenges discussed throughout this paper.

Kuekes, along with electrical engineers, chemists and physicists from around the world, are collaborating with various semiconductor manufacturers and suppliers, government organizations, consortia and universities to promote advancements in the performance of microprocessors and solve some of the challenges that cast doubt on the continuation of the advancements described by Moore's Law.

The mid-term future of microprocessor advancements could very well be based on nanotechnology designs that overcome the physical and quantum problems associated with conventional silicon transistors and processor cores.

Summary Conclusion

The latest advancements in microprocessor technology are well represented within the Intel Itanium Processor, delivering reliability, scalability, security, massive resources, parallelism and a new memory model on a sound microarchitectural foundation.

Because It is so efficient and so small and doesn't depend on out-of-order logic, the latest generation Itanium processor delivers higher performance without creating thermal generation problems. This makes Itanium a simple yet efficient and refined engine that enables more consistent long-term improvement in code execution via small improvements in software, thus reducing the need for significant new advancements in hardware.

Microprocessor hardware improvements are becoming more and more difficult to accomplish as, even Gordon Moore believes, the exponential upward curve in microprocessor hardware advancements "can't continue forever."

* * * * *